



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2013

---

## **Boosting structured additive quantile regression for longitudinal childhood obesity data**

Fenske, Nora ; Fahrmeir, Ludwig ; Hothorn, Torsten ; Rzehak, Peter ; Höhle, Michael

DOI: <https://doi.org/10.1515/ijb-2012-0035>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-87505>

Journal Article

Published Version

Originally published at:

Fenske, Nora; Fahrmeir, Ludwig; Hothorn, Torsten; Rzehak, Peter; Höhle, Michael (2013). Boosting structured additive quantile regression for longitudinal childhood obesity data. *International Journal of Biostatistics*, 9(1):1-18.

DOI: <https://doi.org/10.1515/ijb-2012-0035>

# Nora Fenske\*, Ludwig Fahrmeir, Torsten Hothorn, Peter Rzehak, and Michael Höhle

## Boosting Structured Additive Quantile Regression for Longitudinal Childhood Obesity Data

**Abstract:** Childhood obesity and the investigation of its risk factors has become an important public health issue. Our work is based on and motivated by a German longitudinal study including 2,226 children with up to ten measurements on their body mass index (BMI) and risk factors from birth to the age of 10 years. We introduce boosting of structured additive quantile regression as a novel distribution-free approach for longitudinal quantile regression. The quantile-specific predictors of our model include conventional linear population effects, smooth nonlinear functional effects, varying-coefficient terms, and individual-specific effects, such as intercepts and slopes. Estimation is based on boosting, a computer intensive inference method for highly complex models. We propose a component-wise functional gradient descent boosting algorithm that allows for penalized estimation of the large variety of different effects, particularly leading to individual-specific effects shrunk toward zero. This concept allows us to flexibly estimate the nonlinear age curves of upper quantiles of the BMI distribution, both on population and on individual-specific level, adjusted for further risk factors and to detect age-varying effects of categorical risk factors. Our model approach can be regarded as the quantile regression analog of Gaussian additive mixed models (or structured additive mean regression models), and we compare both model classes with respect to our obesity data.

**Keywords:** longitudinal quantile regression, additive mixed models, body mass index, overweight

**\*Corresponding author: Nora Fenske**, Institut für Statistik, Ludwigs-Maximilians-Universität München, Ludwigstr. 33, München 80539, Germany, E-mail: nora.fenske@stat.uni-muenchen.de

**Ludwig Fahrmeir**, Institut für Statistik, Ludwigs-Maximilians-Universität München, Germany, E-mail: ludwig.fahrmeir@stat.uni-muenchen.de

**Torsten Hothorn**, Abteilung Biostatistik, Universität Zürich, Switzerland, E-mail: torsten.hothorn@uzh.ch

**Peter Rzehak**, Division of Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-Universität München, Germany, E-mail: Peter.Rzehak@med.uni-muenchen.de

**Michael Höhle**, Department of Mathematics, Stockholm University, Sweden, E-mail: hoehle@math.su.se

## 1 Introduction

Obesity is currently considered almost an epidemic and has spread to children during the last decade [1]. Childhood obesity is particularly worrying, because once a child has become obese, it will likely remain obese in adulthood [2]. Therefore, obese children are at high risk for severe long-term sequelae of obesity, such as hypertension, heart diseases, and diabetes mellitus. With the objective of developing effective methods of prevention, enormous public health research efforts have been made to investigate determinants of childhood overweight and obesity [3].

Our work is motivated by and based on the German birth cohort study LISA [4], including data of 2,226 children with up to ten observations each from birth up to the age of 10 years. The aim of our analysis is to flexibly estimate nonlinear age curves of upper quantiles of the body mass index (BMI), both on population and on individual-specific level, while adjusting for early childhood risk factors that have been discussed in the literature [5, 6], such as maternal prenatal lifestyle, parental overweight, socioeconomic factors, and

breastfeeding. In addition, we want to investigate if potential effects of categorical risk factors are time constant or if critical age periods can be identified at which these effects emerge.

This requires to model a quantile function of the form

$$Q_{Y_{ij}}(\tau|\cdot) = h_0(t_{ij}) + \mathbf{x}_{ij}^T \boldsymbol{\beta} + h(v_{ij}) + g(t_{ij}) \cdot u_i + b_{i0} + b_{i1} \cdot t_{ij} \quad [1]$$

for a given quantile parameter  $\tau \in (0, 1)$ . The responses  $y_{ij}$  and covariates are observed for individuals  $i = 1, \dots, N$  at times  $0 \leq t_{i1} < \dots < t_{ij} < \dots < t_{in_i}$ . The predictor has the same structure as in additive mixed models (AMMs) and – more generally – in structured additive regression (STAR) models [7] and contains:

- a smooth nonlinear population trend  $h_0(t)$  where  $t$  denotes child's age
- usual linear population effects  $\boldsymbol{\beta}$  of categorical and continuous covariates  $\mathbf{x}$
- a smooth nonlinear population effect  $h(v)$  of a continuous covariate  $v$ , such as maternal BMI
- an age-varying nonlinear population effect  $g(t)$  of a categorical covariate  $u$ , such as child's sex, and
- individual-specific deviations  $b_{i0} + b_{i1} \cdot t$  from the population trend  $h_0(t)$ , including an individual-specific intercept  $b_{i0}$  and, at least, an individual-specific slope  $b_{i1}$ . Individual-specific age curves are obtained by  $h_0(t) + b_{i0} + b_{i1} \cdot t$ .

For simplicity of presentation, we omitted the quantile parameter  $\tau$  and included only one term of each type in eq. [1], but extensions to several terms are obvious and included in our general model approach in Section 3.2 and in the data analysis.

To the best of our knowledge, few existing models for longitudinal quantile regression cover quantile functions with a structured additive predictor as in eq. [1]. The majority of existing approaches considers predictors with conventional linear population effects and individual-specific intercepts. They either rely on a penalized loss criterion where shrinkage of individual-specific effects to zero is imposed [8], on quasi-likelihood approaches based on the asymmetric Laplace error distribution as a working likelihood [9–11], or on full Bayesian inference with a nonparametric Dirichlet process error distribution [12]. On the other hand, additive quantile regression models without individual-specific effects have been recently proposed by Fenske et al. [13], using boosting as computationally effective method for smoothing, and by Koenker [14, 15], using a total variation penalty for enforcing smoothness of the functional effects.

The quantile function of the model in Yue and Rue [16] comes closest to structured additive quantile predictors as in eq. [1], although it contains only a random intercept and no varying-coefficient terms. A certain drawback of this Bayesian approach is the assumption of an asymmetric Laplace error distribution, which is only a working model and possibly too restrictive for adequately approximating the true error distribution. Thus, it is somewhat unclear how well (quasi-)posteriors are appropriate for inferential purposes in general.

We, therefore, decided to extend the boosting approach for additive quantile regression in Fenske et al. [13] to structured additive quantile regression (STAQ) for longitudinal data. Boosting is a distribution-free inference method, that is, it does not require distributional assumptions on the errors, and estimation is based on a loss function, which is the check function in case of quantile regression. Component-wise functional gradient descent boosting automatically provides smoothing of nonlinear effects, shrinkage of individual-specific effects, and an implicit possibility for variable selection. For inferential purposes, such as providing standard errors of the estimates, we apply block-wise bootstrapping.

Conceptually, the additive model in Koenker [15] could also be extended to longitudinal quantile regression by including a  $L_1$ -norm penalty for shrinking the individual-specific effects. However, the computational effort for solving the resulting linear program will probably pose a serious challenge.

Our analysis is also innovative from an epidemiological point of view. A typical statistical approach for analyzing childhood overweight and obesity would be to classify children as obese using reference charts, followed by multiple logistic regression for the resulting binary response [6, 17]. Since the BMI distribution typically becomes right-skewed with increasing age, the construction of such reference charts is a challenging task and has been tackled by various statistical approaches in the past, see Borghi et al. [18] for an

overview. In our investigation, in contrast, we directly model upper BMI quantiles of the study population, such as the 90% BMI quantile for overweight and the 97% BMI quantile for obesity. Therefore, we avoid possible loss of information implied by reducing the original continuous response BMI to the binary response obesity. Since logit and probit models assume a specific symmetric distribution for the original continuous response variable, the age-specific skewness of BMI distributions makes the use of conventional binary regression models questionable.

For cross-sectional BMI data, quantile regression methods have been used to model a Z-score of the BMI depending on covariates [19, 20], which was obtained by transforming raw BMI values based on age- and sex-specific reference charts. Here, we directly model raw BMI quantiles and include age and sex as covariates; and we thereby avoid the decision for a specific reference chart.

In our analysis, we compare structured additive median regression with Gaussian AMMs which can be regarded as a special instance of STAR models for longitudinal data. AMMs currently provide the only possibility to fit the full variety of population and individual-specific effects of eq. [1] and are, thus, the only serious competitor of our model. AMMs not only imply conditional mean modeling but can also be used for quantile regression, since the conditional response distribution is completely determined by the *iid* Gaussian assumption for the error terms. Thus, AMMs are an appropriate approach for quantile regression when the distribution of the conditional response is homoscedastic and approximately Gaussian. We would like to stress this fact, because one can then benefit from AMMs as a well-studied, established, and implemented framework for STAQ models for longitudinal data, as illustrated in a pre-analysis of the LISA study [21].

## 2 Data description

The LISA study is a large prospective longitudinal birth cohort study in four German cities (Bad Honnef, Leipzig, Munich, and Wesel), in which 3,097 healthy neonates born between 11/1997 and 01/1999 were originally included. The follow-up time was until the age of 10 years, and data were collected through questionnaires covering the nine mandatory well-baby check-up examinations by a pediatrician at birth and the age of 2 weeks and 1, 3, 6, 12, 24, 48, and 60 months. For the 10-year (120 months) follow-up, anthropometric measurements were taken by physical examination at the study centers. Thus, the maximum number of observations per child was 10. Further details on the LISA study can be found in Breitfelder et al. [22] and Rzehak et al. [23].

In our analysis, we followed a complete case approach. When an observation of a time-constant covariate was missing, we excluded all observations of the respective child from the analysis. When, on the other hand, a single observation of age or BMI was missing, only this particular observation of the respective child was excluded from the analysis. Altogether, a total of 19,819 observations of 2,226 children were included in the analyses and for statistical modeling. The decision for the complete case approach resulted from several analyses with respect to missing data and dropout, suggesting the missing data mechanism to be “missing at random”.

Tables 1 and 2 give an overview of the continuous and categorical variables, respectively, included in the analysis.

**Table 1** Description of continuous variables for complete cases.

Variable	Abbrev.	Unit	Median	Mean	SD	N
<i>Time-varying variables</i>						
BMI	BMI	kg/m <sup>2</sup>	15.36	15.28	2.08	19,819
Age	Age	Years	0.54	1.86	2.64	19,819
<i>Time-constant variables</i>						
Maternal BMI at pregnancy begin	mBMI	kg/m <sup>2</sup>	21.72	22.59	3.76	2,226
Maternal BMI gain during pregnancy	mBMlgain	kg/m <sup>2</sup>	4.95	5.12	1.67	2,226

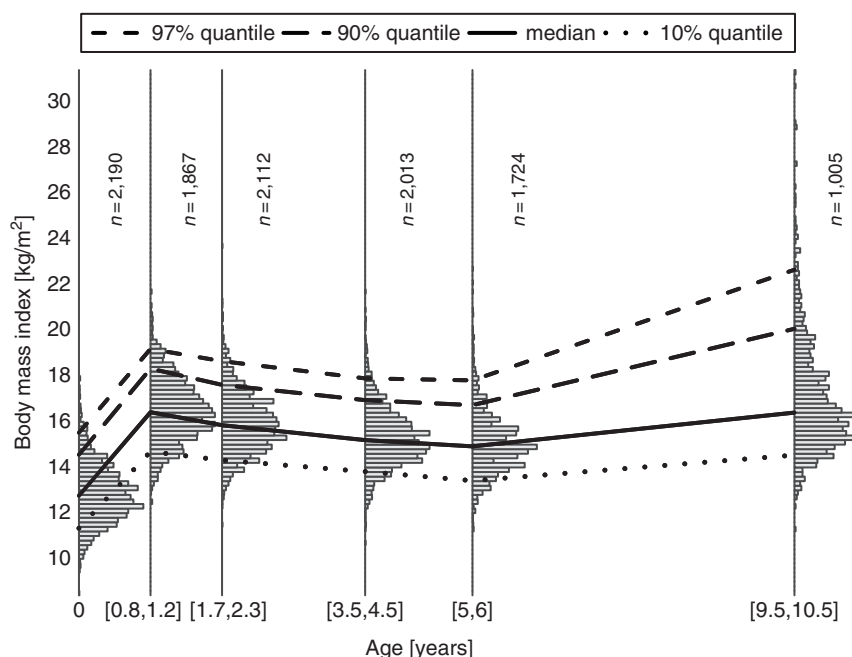
**Table 2** Description of categorical variables.

Covariate	Abbrev.	Categories	Frequ. in (%)	<i>N</i>
Sex	Sex	0 = Female	47.8	1,064
		1 = Male	52.2	1,162
Study location	Location	0 = Rural (Bad Honnef, Wesel)	21.5	478
		1 = Urban (Leipzig, Munich)	78.5	1,748
Nutrition until the age of 4 months	Nutri	0 = Bottle fed and/or breast fed	41.2	917
		1 = Breast fed only	58.8	1,309
Maternal smoking during pregnancy	Smoke	0 = No	85.0	1,899
		1 = Yes	15.0	327
Maternal highest level of education	mEdu	1 = Certificate of secondary education (CSE) or lower-level secondary school (Hauptschule)	7.0	157
		2 = Secondary school (Realschule)	35.8	798
		3 = High school (Abitur/Fachabitur)	57.1	1,271

Note: Absolute frequencies *N* relate to complete cases of 2,226 children.

The covariates cover most of the early childhood risk factors that are discussed in the literature, such as parental overweight (maternal BMI), socioeconomic factors (urban/rural location and maternal education), nutrition (breastfeeding), and maternal prenatal lifestyle factors, such as maternal BMI gain during pregnancy and maternal prenatal smoking which are believed to be associated with rapid postnatal growth of the offspring [24–26]. All variables except for age and BMI are time constant.

Figure 1 displays scale and skewness of the empirical BMI distribution by age. Thereby, the relationship between age and skewness of the empirical BMI distribution can be inspected, which is an important tool



**Figure 1** Empirical BMI distribution by age. Relative frequencies in the histograms were calculated based on all observations within an age interval. The six age intervals are shown on the x-axis, each *n* denotes the total number of observations in the interval. Vertical lines are drawn at the midpoints of the intervals. Lines connecting the age-specific empirical 10%, 50%, 90%, and 97% quantiles of the BMI distribution within each class are also shown.

for checking the distributional assumptions for the later modeling, that is, a Gaussian distribution conditional on covariates in the case of AMMs. Figure 1 suggests an age-specific skewness of the BMI distribution, beginning after the age of 6 years. Note that no adjustment for additional covariate information is done in this plot. A main goal of the present analysis is to model nonlinear population and individual-specific age curves of upper BMI quantiles that adequately reflect the shape of the BMI distribution in Figure 1, while adjusting for covariates other than age. In addition, it is of interest if the shape of the BMI distribution changes for different levels of the categorical covariates.

### 3 Statistical modeling

This section briefly explains Gaussian AMMs with respect to quantile regression and introduces the general model class of STAQ for longitudinal data, followed by a description of the boosting algorithm and of the specific model used for our analysis.

Throughout the section, we use the following notation: (i) indices  $i = 1, \dots, N$ , for individuals and  $j = 1, \dots, n_i$ , for intra-individual observations; (ii) total number of observations  $n = \sum_{i=1}^N n_i$ ; (iii) response variable  $y_{ij}$  for individual  $i$  at observation  $j$ , later corresponding to the BMI, design vectors  $\mathbf{x}_{ij} = (1, x_{ij1}, \dots, x_{ijp})^T$  for  $p$  covariates with linear effects, and  $\mathbf{v}_{ij} = (v_{ij1}, \dots, v_{ijm})^T$  for  $m$  continuous covariates with potentially nonlinear effects on the response; (iv) quantile function  $Q_Y(\tau|\mathbf{x}) = F_Y^{-1}(\tau|\mathbf{x})$  for the  $\tau \cdot 100\%$  quantile with  $\tau \in (0, 1)$  of the response variable  $Y$  conditional on a given covariate vector  $\mathbf{x}$ , where  $F_Y^{-1}(\tau|\mathbf{x})$  denotes the inverse cdf of  $Y|\mathbf{x}$ .

#### 3.1 Gaussian additive mixed models

Linear and AMMs are a common statistical approach to model relationships between covariates and the conditional mean of a response variable in longitudinal data; see, for example, Ruppert et al. [27]. The AMM considered here can be written as

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + h_1(v_{ij1}) + \dots + h_m(v_{ijm}) + \mathbf{z}_{ij}^T \mathbf{b}_i + \varepsilon_{ij} = \eta_{ij}^{(\mu)} + \mathbf{z}_{ij}^T \mathbf{b}_i + \varepsilon_{ij} \quad [2]$$

with *iid* errors  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ . The linear part  $\mathbf{x}_{ij}^T \boldsymbol{\beta}$  of the additive predictor  $\eta_{ij}^{(\mu)}$  contains usual linear population effects  $\boldsymbol{\beta}$  of covariates, while the unknown functions  $h_l(\cdot)$ ,  $l = 1, \dots, m$ , describe possible smooth nonlinear population effects of the basic time scale and of further continuous covariates. This population part of the predictor is denoted as  $\eta_{ij}^{(\mu)}$  and can also be extended to varying-coefficient terms  $u_{i1} \cdot g_1(\text{age}_{ij}) + \dots + u_{ir} \cdot g_r(\text{age}_{ij})$ , where  $u_{i1}, \dots, u_{ir}$  are  $r$  time-constant categorical covariates (subvectors of  $\mathbf{x}_{ij}$ ) and  $g_1(\text{age}_{ij}), \dots, g_r(\text{age}_{ij})$  denote age-varying effects that are estimated in analogy to  $h_l(\cdot)$ . The random effects  $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$  are independent for different individuals  $i$ , and also independent from the errors  $\varepsilon_{ij}$ . By including time-varying covariates such as age in the design vector  $\mathbf{z}_{ij}$ , model (2) allows for the estimation of individual-specific random slopes or curves. Thus, the model accounts for the correlation between repeated intra-individual measurements and thereby avoids confounding of the covariate effects with latent heterogeneity between individuals. When only a random intercept  $b_{i0}$  is included in the model equation, for example, an equicorrelation between intra-individual response observations is induced.

In the case of Gaussian distributed error terms, AMMs not only model the mean but also the quantile function of the response's distribution and thereby imply quantile regression. When the resulting quantile function of the response variable is regarded, we have to distinguish between

the conditional quantile function:  $Q_{Y_{ij}}(\tau|\eta_{ij}^{(\mu)}, \mathbf{b}_i) = \eta_{ij}^{(\mu)} + \mathbf{z}_{ij}^T \mathbf{b}_i + q_\tau \sigma_\varepsilon$

and the marginal quantile function  $Q_{Y_{ij}}(\tau|\eta_{ij}^{(\mu)}) = \eta_{ij}^{(\mu)} + q_\tau \sqrt{\mathbf{z}_{ij}^T \boldsymbol{\Sigma}_b \mathbf{z}_{ij} + \sigma_\varepsilon^2}$

where  $q_\tau$  denotes the  $\tau \cdot 100\%$  quantile of a standard Gaussian distribution. For covariates that are just contained in the predictor  $\eta_{ij}^{(\mu)}$ , the interpretation of covariate effects with respect to the quantile functions remains the same as for the mean. In the case of time-varying covariates included in  $\mathbf{z}_{ij}$ , however, the relationship between covariate and quantile functions becomes more involved, since both quantile functions depend on the design of  $\mathbf{z}_{ij}$ . When only a random intercept is included in the model equation, both quantile functions reduce to a simple time-constant shift of the population predictor  $\eta_{ij}^{(\mu)}$ .

However, AMMs are not adequate for quantile regression if higher moments (variance, skewness, or kurtosis) of the conditional response's distribution depend on covariates, which means that the *iid* Gaussian error assumption is violated. As shall be shown, this will be the case for the LISA study analyzed here because of an age-specific skewness of the BMI distribution.

With regard to estimation, AMMs are a well-studied and established framework for longitudinal mean regression, and estimation algorithms and software are highly developed. Common approaches for estimating AMMs rely on penalized likelihood concepts, see, for example, Ruppert et al. [27] and are aimed at estimating smooth nonlinear covariate effects based on penalized spline functions [28]. For fitting AMMs in our analysis, we used the R package *amer* [29]. This package has recently been retired from the CRAN repository, since its scope of functionality is completely covered by the more sophisticated R package *gamm4* [30].

### 3.2 Structured additive quantile regression for longitudinal data

Distribution-free quantile regression models are directed at modeling specific  $\tau \cdot 100\%$  quantiles of a response variable without assuming any distribution for response or errors. The general approach of quantile regression is thoroughly treated in Koenker [14]. For a fixed value of  $\tau$ , the STAQ model for longitudinal data can be written as:

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_\tau + h_{\tau 1}(\mathbf{v}_{ij1}) + \cdots + h_{\tau m_\tau}(\mathbf{v}_{ijm_\tau}) + \mathbf{z}_{ij}^T \mathbf{b}_{\tau i} + \varepsilon_{\tau ij} \quad [3]$$

$$= \eta_{ij}^{(\tau)} + \mathbf{z}_{ij}^T \mathbf{b}_{\tau i} + \varepsilon_{\tau ij} \quad \text{where} \quad \varepsilon_{\tau ij} \sim F_{\varepsilon_{\tau ij}} \quad \text{and} \quad F_{\varepsilon_{\tau ij}}(0) = \tau. \quad [4]$$

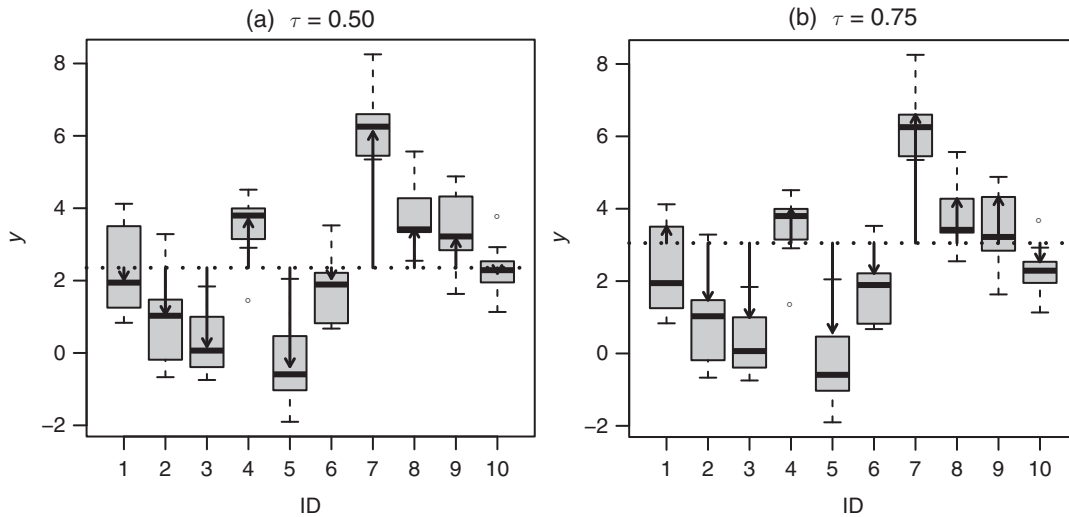
At first glance, the predictor looks similar to the predictor in AMMs. However, the linear effects  $\boldsymbol{\beta}_\tau$ , the functions  $h_{\tau l}(\cdot)$ ,  $l = 1, \dots, m_\tau$ , and the individual-specific effects  $\mathbf{b}_{\tau i}$  are now quantile specific. Also, the design vectors  $\mathbf{x}_{ij}$  and  $\mathbf{z}_{ij}$  may depend on the given quantile, meaning that for different quantile parameters different covariates can be included in the predictor. The error terms  $\varepsilon_{\tau ij}$  in eq. [4] are assumed to be mutually independent, but no specific distributional assumption is made for them apart from the restriction  $F_{\varepsilon_{\tau ij}}(0) = \tau$ . A consequence of this condition is that model (3) implies quantile regression of the response conditional on covariates and individual-specific effects, that is, the conditional quantile function has an additive structure

$$Q_{Y_{ij}}(\tau|\eta_{ij}^{(\tau)}, \mathbf{b}_{\tau i}) = q_{\tau ij} = \eta_{ij}^{(\tau)} + \mathbf{z}_{ij}^T \mathbf{b}_{\tau i}. \quad [5]$$

To ease later notation, the complete predictor is denoted by  $q_{\tau ij}$ . We assume the individual-specific effects  $\mathbf{b}_{\tau i}$  to be independent for different  $i$  but without a specific distribution to conserve the distribution-free character of the model. As shall be seen in Section 3.3, we will estimate the  $\mathbf{b}_{\tau i}$  with a ridge-type penalty



which leads to shrunken individual-specific effects. Due to the quadratic form of the penalty term (corresponding to the log-density of Gaussian random effects priors from a Bayesian perspective), the individual-specific effects can be interpreted in accordance with the conditional view of AMMs. Similar to individual-specific conditional means scattering around the population mean in AMMs, individual-specific conditional  $\tau \cdot 100\%$  quantiles scatter symmetrically around the  $\tau \cdot 100\%$  population quantile in STAQ models for longitudinal data. To illustrate this interpretation, Figure 2 shows the results of a small simulation example.



**Figure 2** Boxplots displaying empirical distributions of simulated data for ten individuals with ten observations each. (a) Results from an STAQ model for  $\tau = 0.50$  containing only a population intercept and individual-specific intercepts. (b) Results from a similar STAQ model for  $\tau = 0.75$ . Dotted horizontal lines correspond to the estimated respective population quantiles, whereas individual vertical arrows stand for estimated individual-specific intercepts as deviations from the population quantile.

In both plots, boxplots display the empirical distributions of the same data simulated from the model  $y_{ij} = \beta_0 + b_i + \varepsilon_{ij}$  with  $i = 1, \dots, 10$ , individuals and  $j = 1, \dots, 10$ , observations per individual, where  $b_i$  and  $\varepsilon_{ij}$  were both drawn from a Gaussian distribution with variances of 4 and 1, respectively. We fitted two STAQ models by boosting for  $\tau = 0.50$  and  $\tau = 0.75$ , containing a population intercept as well as quantile- and individual-specific intercepts. The results for the median are shown in panel (a), while panel (b) shows the results for the 75% quantile. It can be seen that the individual-specific intercepts differ for different quantiles. Also, the sums of the estimated population effects and individual-specific intercepts roughly correspond to the respective individual-specific empirical quantiles given by the boxplots. Extensions of the model to covariates and individual-specific slopes would not change the interpretation of individual-specific effects in longitudinal quantile regression. This illustration additionally points out that the interpretation of population effects is conditional on individual-specific effects, corresponding to the conditional quantile function of AMMs.

### 3.3 Boosting estimation for STAQ models

In general, estimation of distribution-free quantile regression models relies on the minimization of an empirical loss criterion, which is in accordance with eq. [5] given as



$$\sum_{i=1}^N \sum_{j=1}^{n_i} \rho_{\tau}(y_{ij}, q_{\tau ij}), \quad [6]$$

where the *check function*

$$\rho_{\tau}(y, q_{\tau}) = \begin{cases} \tau \cdot |y - q_{\tau}| & y \geq q_{\tau} \\ (1 - \tau) \cdot |y - q_{\tau}| & y < q_{\tau} \end{cases}$$

is the standard loss function for quantile regression models and is proportional to the absolute value loss for  $\tau = 0.5$ , i.e.  $\rho_{0.5}(y, q_{0.5}) = 0.5|y - q_{0.5}|$ . Standard approaches for solving the optimization problem in eq. [6] rely on linear programming [8, 14].

For fitting STAQ models, we use a component-wise functional gradient descent boosting algorithm to minimize the criterion in eq. [6], as has been recently proposed for estimating additive quantile regression models [13]. In brief, component-wise functional gradient descent boosting is an optimization algorithm that aims at minimizing an empirical loss criterion as given in eq. [6] by stepwise updating an estimator according to the steepest gradient descent of the loss criterion; see Bühlmann and Hothorn [31] for a general overview. In addition to smooth nonlinear and time-varying effects already explored [13], we here also include individual-specific effects in the additive predictor and estimate them with penalized least-squares base learners. In the following, we briefly describe the adapted boosting algorithm for estimating STAQ models for longitudinal data. Our description is based on the general boosting algorithm treated in Fenske et al. [13], and we here only concentrate on the adaptations that were necessary to make the penalized estimation of individual-specific effects possible. To ease the notation, we suppress the index  $\tau$ .

After having initialized the model parameters contained in the predictor  $q_{ij}$  with suitable starting values, the negative gradient residuals of the empirical risk in iteration  $m = 1, \dots, M$  can be expressed as

$$u_{ij}^{[m]} = \begin{cases} \tau & y_{ij} \geq \hat{q}_{ij}^{[m-1]} \\ \tau - 1 & y_{ij} < \hat{q}_{ij}^{[m-1]} \end{cases},$$

where  $\hat{q}_{ij}^{[m-1]}$  represents the estimated predictor from iteration  $m - 1$ . In every iteration  $m$  of the boosting algorithm, the negative gradient residuals  $u_{ij}^{[m]}$  are used as working responses for the base learners, which can be described as univariate regression models and are fitted separately for each covariate. For linear population effects  $\beta$ , base learners typically correspond to simple univariate least-squares models. In the case of smooth functional effects  $h_l(\cdot)$ , base learners correspond to penalized least-squares models estimated by P-splines. Here, we only present the specific base learners for estimating individual-specific effects.

First of all, the individual-specific effects in  $\mathbf{b}_i = (b_{i0}, b_{i1}, \dots, b_{iK})^T$  from the STAQ model in eq. [5] are separated into  $K + 1$  different base learners. In analogy to Gaussian random effects in AMMs, a natural concept for estimating individual-specific effects is penalized least-squares base learners with ridge-type penalties. Thus, for fitting the vectors of  $N$  individual-specific effects  $\mathbf{b}_k = (b_{1k}, \dots, b_{Nk})^T$  for  $k = 0, \dots, K$ , we denote the corresponding  $k$ th base learner with  $\mathbf{a}_k = (a_{1k}, \dots, a_{Nk})^T$ . We derive the penalized least-squares optimization criterion to be minimized as

$$\sum_{i=1}^N \sum_{j=1}^{n_i} (u_{ij}^{[m]} - a_{ik})^2 + \lambda \sum_{i=1}^N a_{ik}^2$$

with smoothing parameter  $\lambda$  controlling the degree of shrinkage of the individual-specific effects. The fitted base learner can then be written as

$$\hat{\mathbf{a}}_k^{[m]} = (\mathbf{Z}_k^T \mathbf{Z}_k + \lambda \mathbf{I}_N)^{-1} \mathbf{Z}_k^T \mathbf{u}^{[m]},$$

where the  $(n \times 1)$ -vector  $\mathbf{u}^{[m]}$  contains negative gradient residuals of all observations at iteration  $m$ ,  $\mathbf{I}_N$  represents the identity matrix with dimension  $N$ , and  $\mathbf{Z}_k$  is a  $(n \times N)$  design matrix and depends on the type of base learner. The fitted negative gradient residuals can then be expressed as  $\hat{\mathbf{u}}^{[m]} = \mathbf{S}_\lambda \mathbf{u}^{[m]}$  with smoother matrix  $\mathbf{S}_\lambda = \mathbf{Z}_k (\mathbf{Z}_k^T \mathbf{Z}_k + \lambda \mathbf{I}_N)^{-1} \mathbf{Z}_k^T$ .

For fitting individual-specific intercepts  $\mathbf{b}_0 = (b_{10}, \dots, b_{N0})^T$ , the matrix  $\mathbf{Z}_0$  is a zero-one indicator matrix, such that the  $(n \times 1)$ -vector resulting from  $\mathbf{Z}_0 \mathbf{a}_0$  contains respective individual-specific intercepts. For fitting individual-specific slopes  $\mathbf{b}_1 = (b_{11}, \dots, b_{N1})^T$  of a time-varying covariate  $z$ , the design matrix  $\mathbf{Z}_1$  links individual-specific slopes to the corresponding observations of covariate  $z$ , leading to:

$$\mathbf{Z}_0 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \end{pmatrix} \quad \mathbf{Z}_1 = \begin{pmatrix} z_{11} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ z_{1n_1} & 0 & \cdots & \cdots & 0 \\ 0 & z_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & z_{2n_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \end{pmatrix}.$$

To summarize, in every boosting iteration  $m$ , all possible base learners are fitted, but only the parameters corresponding to the best-fitting base-learning procedure are updated, that is, the base learner that minimizes the squared error loss between fitted values and negative gradient residuals is updated. Therefore, the boosting algorithm has a component-wise character and provides an inherent variable selection property, since some covariates will be chosen more often during the boosting process and earlier for the first time than others. For example, when the individual-specific intercept is the best-fitting base-learning procedure, its corresponding update is given as

$$\hat{b}_{i0}^{[m]} = \hat{b}_{i0}^{[m-1]} + \nu \cdot \hat{a}_{i0}^{[m]} \quad i = 1, \dots, N.$$

The parameter  $\nu \in (0, 1]$  controls the step length and is implicitly connected to the total number of boosting iterations  $m_{\text{stop}}$ . The smaller the  $\nu$  chosen the larger will be the optimal  $m_{\text{stop}}$ , which can, for example, be determined by cross-validation of the empirical risk. Early stopping is usually done to avoid overfitting and leads to biased estimates shrunken toward zero [31]. This is also the reason why the individual-specific effects of the STAQ model can be regarded as shrunken fixed effects.

Note that with boosting, the smoothing parameter  $\lambda$  is fixed in advance and not considered as a hyperparameter to be optimized [32]. Instead of choosing  $\lambda$  based on generalized cross-validation or similar criteria as in classical AMM estimation, the degree of smoothness of a penalized effect is controlled by the number of boosting iterations  $m_{\text{stop}}$ . Due to the repeated selection of a base learner, the final degree of smoothness can still achieve a higher order than the one imposed by the initial degrees of freedom [31]. In addition, the initial degrees of freedom should be fixed at the same (small) value for all base learners to ensure that the complexity and selection probability of each base learner is comparable [32].

With regard to software, boosting for quantile regression can be embedded in the well-studied and implemented class of component-wise functional gradient descent boosting algorithms. We used the

function `gamboost` from package `mboost` [33, 34] with options `family=QuantReg()` and base learner `brandom()`; see the Appendix for details on the model specification, implementation, and estimation.

### 3.4 Longitudinal STAQ model for the LISA study

For our analysis of the LISA study, we estimated STAQ models for the 90% and 97% BMI quantiles and – for reasons of model comparison – for the median and the 10% quantile. We considered the following predictor for  $\tau \in \{0.10, 0.50, 0.90, 0.97\}$ :

$$\text{BMI}_{ij} = \eta_{ij}^{(\tau)} + b_{\tau i0} + b_{\tau i1} \cdot \text{Age}_{ij} + \varepsilon_{\tau ij}, \quad [7]$$

where the population part is given by

$$\begin{aligned} \eta_{ij}^{(\tau)} = & \beta_0 + h_{\tau \text{Age}}(\text{Age}_{ij}) + h_{\tau \text{mBMI}}(\text{mBMI}_i) + h_{\tau \text{mBMIgain}}(\text{mBMIgain}_i) \\ & + \beta_{\tau \text{Sex}} \text{Sex}_i + \beta_{\tau \text{Location}} \text{Location}_i + \beta_{\tau \text{Nutri}} \text{Nutri}_i + \beta_{\tau \text{Smoke}} \text{Smoke}_i \\ & + \beta_{\tau \text{mEdu2}} \text{mEdu2}_i + \beta_{\tau \text{mEdu3}} \text{mEdu3}_i \\ & + \text{Sex}_i \cdot g_{\tau \text{Male}}(\text{Age}_{ij}) + \text{Location}_i \cdot g_{\tau \text{Urban}}(\text{Age}_{ij}) + \text{Nutri}_i \cdot g_{\tau \text{Nutri}}(\text{Age}_{ij}) \\ & + \text{Smoke}_i \cdot g_{\tau \text{Smoke}}(\text{Age}_{ij}) + \text{mEdu2}_i \cdot g_{\tau \text{mEdu2}}(\text{Age}_{ij}) + \text{mEdu3}_i \cdot g_{\tau \text{mEdu3}}(\text{Age}_{ij}). \end{aligned}$$

Our model thus contains main effects for the entire set of covariates given in Tables 1 and 2, and age-varying effects of categorical covariates. For the boosting algorithm, we defined separate base learners for all smooth effects. All continuous covariates, including age and age-varying effects, were modeled by penalized least-squares base learners based on P-splines with  $\text{df}(\lambda) = 5$ . The three-level covariate `mEdu` was split into two dummy-coded variables relating to the reference category “No degree or CSE”, but was fitted in one single base learner with  $\text{df}(\lambda) = 5$  together with main effects of all other categorical covariates. To account for the longitudinal data structure, we included an individual-specific intercept  $b_{\tau i0}$  as well as a slope  $b_{\tau i1}$  for age in the model, describing individual-specific deviations from the population effect for age. These effects were again fitted by penalized least-squares base learners with  $\text{df}(\lambda) = 5$  to equalize selection probabilities of different base learners as described above. The number of boosting iterations  $m_{\text{stop}}$  was chosen based on block-wise fivefold cross-validation and resulted in about 5,000 iterations per model. We set the step length  $\nu = 0.4$ , since this results in fewer boosting iterations and therefore lower computational effort than for smaller values of  $\nu$ . For reasons of model comparison, we estimated an AMM with the same predictor as in eq. [7].

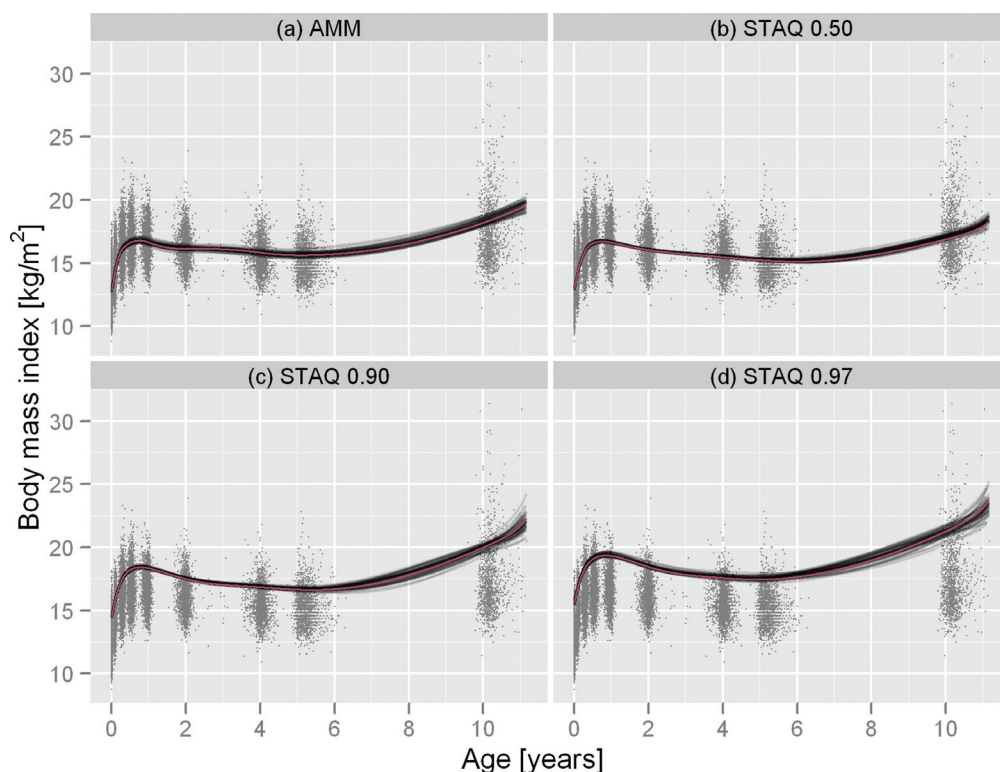
Since boosting does not directly provide standard errors for the estimated effects, we additionally conducted a block-wise bootstrap analysis in accordance with Liu and Bottai [10]. We obtained one single bootstrap sample by randomly choosing 2,226 children with replacement at the first stage. To conserve the longitudinal data structure, all observations corresponding to the chosen children were included in the bootstrap sample at the second stage. In this way, we generated a total of 50 different bootstrap samples and used each sample to fit STAQ models and AMMs as described above.

To formally compare the results from AMMs and STAQ models, we additionally constructed 50 out-of-bag samples with children that were not contained in the respective bootstrap samples. These out-of-bag samples were used to calculate the empirical risks based on the check function as given in eq. [6] for the three different quantiles and two different model classes. To obtain an estimated predictor  $\hat{q}_{\tau ij}$  for a child in an out-of-bag sample, we set its individual-specific effects to zero in order to obtain the empirical risk for model comparison.

In a simultaneous analysis with respect to missing data, we created an imputed version of the data using several missing data imputation methods. Then, we repeated all statistical analyses with the imputed data and obtained similar results as from the complete case approach.

## 4 Results

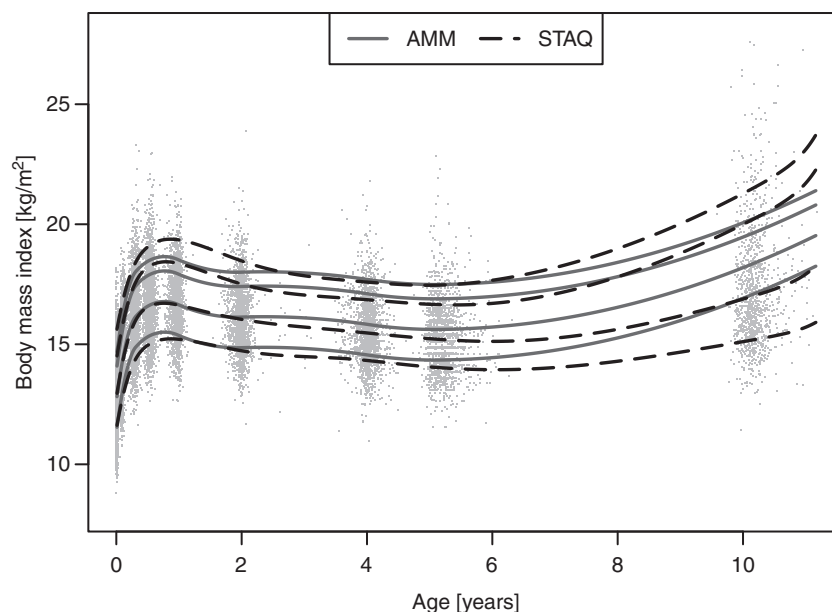
The resulting smooth nonlinear population effects of age on BMI quantiles are shown in Figure 3. Overall, the shape of the age effect remains stable over the bootstrap iterations for all models and confirms the first impression from the descriptive analysis in Figure 1. In sparse data regions, that is, between the ages of 6 and 10 years, the variation of the effects is larger than in regions with more observations. The effects for the AMM and STAQ median look roughly similar. For upper quantiles, the age effect strongly increases beginning after the age of 6 years.



**Figure 3** Estimated nonlinear BMI quantile curves by age resulting from (a) AMM and from STAQ models for (b)  $\tau = 0.50$ , (c)  $\tau = 0.90$ , and (d)  $\tau = 0.97$ . BMI observations (in gray) with estimated curves (in black) for 50 different bootstrap samples are shown; the superimposed red line represents the estimated nonlinear quantile curve of the respective model on the full dataset. Quantile curves are adjusted for other covariates (fixed combination).

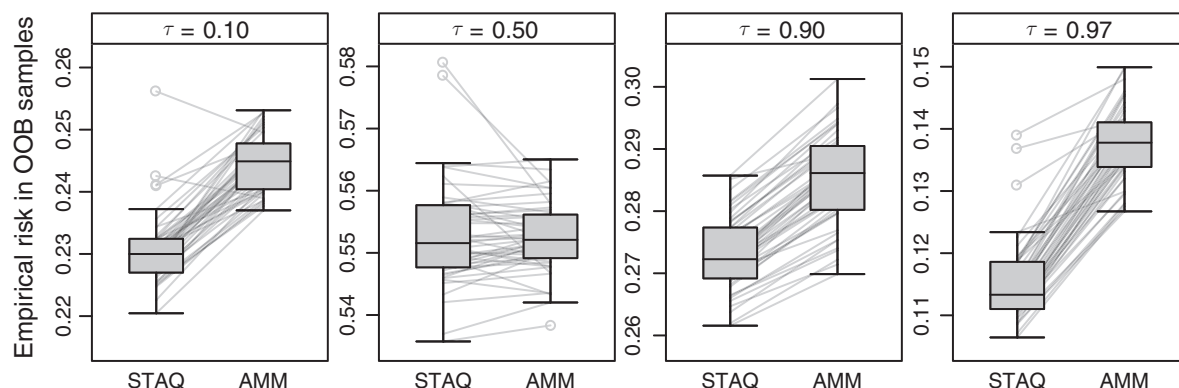
Furthermore, Figure 4 shows estimated age-specific quantile curves from the two model classes. To make the effects comparable, we concentrate on the population quantile functions conditional on individual-specific effects. Thus, AMM curves for upper quantiles were obtained by a parallel shift of the median curve, whereas for STAQ models, all quantile curves were modeled separately. The resulting curves are estimated to be roughly similar until the age of 6 years. At the age of 10 years, the 90% BMI curve estimated by STAQ regression is above the 97% curve estimated by AMMs, whereas the 10% STAQ curve is below the AMM median.

We additionally compared the model fits by block-wise bootstrap and calculated the empirical quantile-specific risks based on the check function in the 50 out-of-bag bootstrap samples, as described in



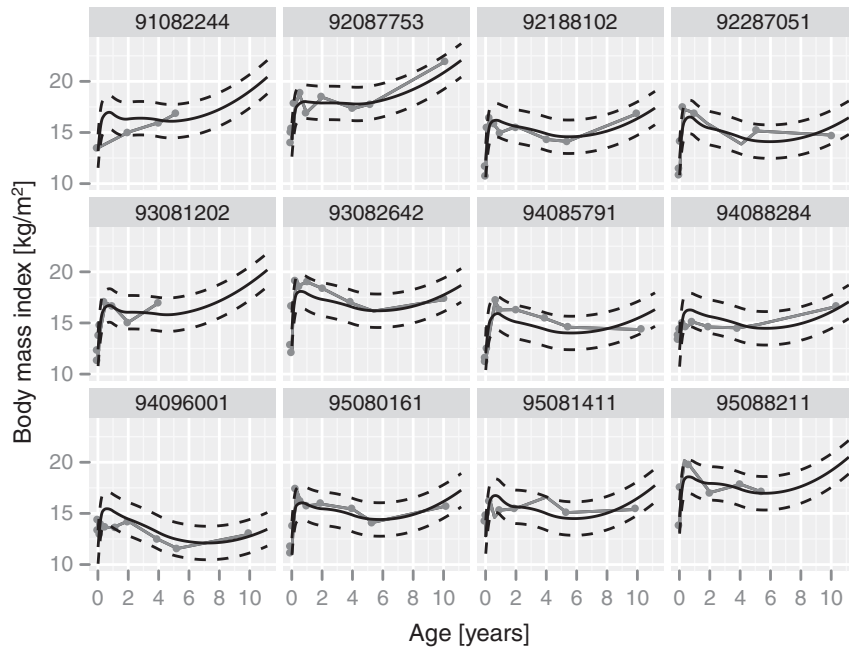
**Figure 4** Comparison of AMM and STAQ results for the four quantiles  $\tau = 0.10$ ,  $\tau = 0.50$ ,  $\tau = 0.90$ , and  $\tau = 0.97$ . BMI observations (in gray) with estimated nonlinear quantile curves for the BMI depending on age and a fixed covariate combination of time-constant covariates are shown.

Section 3.4. Figure 5 shows that there are no fundamental differences between the empirical risks for the median, but that STAQ models clearly outperform AMMs for other quantiles. This result is in accordance with Figure 4, which together demonstrates that STAQ models are more appropriate for handling the age-specific skewness of the BMI distribution than AMMs.

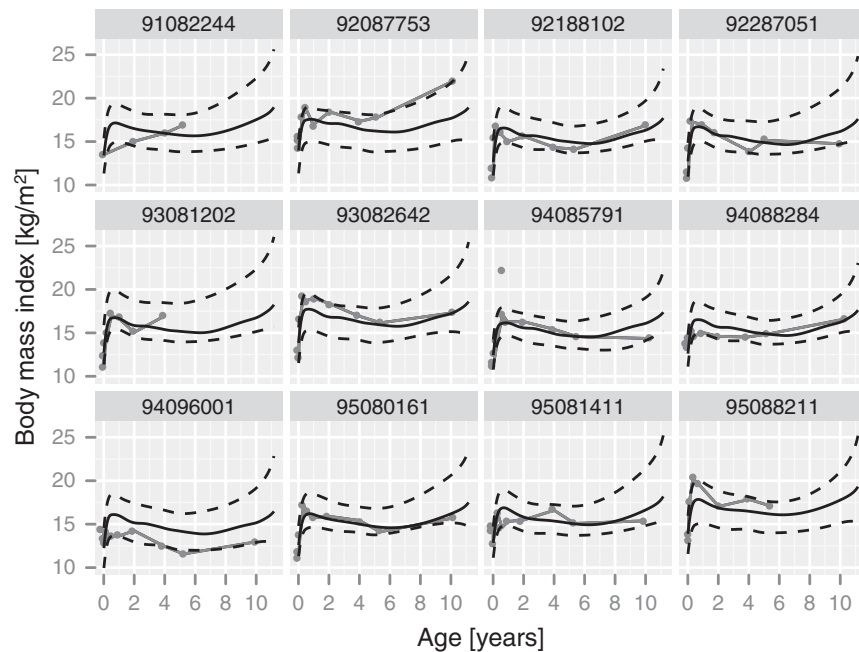


**Figure 5** Risk comparison of STAQ and AMM for the four quantiles  $\tau = 0.10$ ,  $\tau = 0.50$ ,  $\tau = 0.90$ , and  $\tau = 0.97$ . Boxplots show empirical distributions of the empirical risks calculated on 50 out-of-bag (OOB) bootstrap samples. Results for one out-of-bag sample are connected by gray lines.

To assess the uncertainty for individual-specific mean predictions, we constructed individual-specific 90% prediction intervals based on estimated 5% and 95% BMI quantile curves. This procedure was based on the suggestion in Meinshausen [35]. Figure 6 shows individual-specific BMI quantile curves depending on age for 12 randomly chosen children estimated by AMMs. The dashed quantile curves, corresponding to the interval limits, are parallel shifts of the mean curves. The symmetric shape and the distance between the



**Figure 6** Individual-specific BMI quantile curves estimated by AMMs for 12 randomly chosen children. Continuous black line shows the estimated median, while the dashed lines show the estimated 5% and 95% quantile curves, respectively. Observed BMI values are displayed by gray line-connected points.



**Figure 7** Individual-specific BMI quantile curves estimated by STAQ models for 12 randomly chosen children, in analogy to Figure 6.

curves remain the same for all children. The offset differences between children can be attributed to the child-specific intercept and covariate combination; the shape differences can be attributed to the child-specific slopes. One can see that the mean curve reproduces the true BMI pattern (in gray) in most cases.

For STAQ models, individual-specific BMI quantile curves are shown in Figure 7. Since the three quantile curves are estimated independently from each other, the quantile curves are no longer parallel shifts of the median curves, as was the case for AMMs. The interval widths vary notably between children,

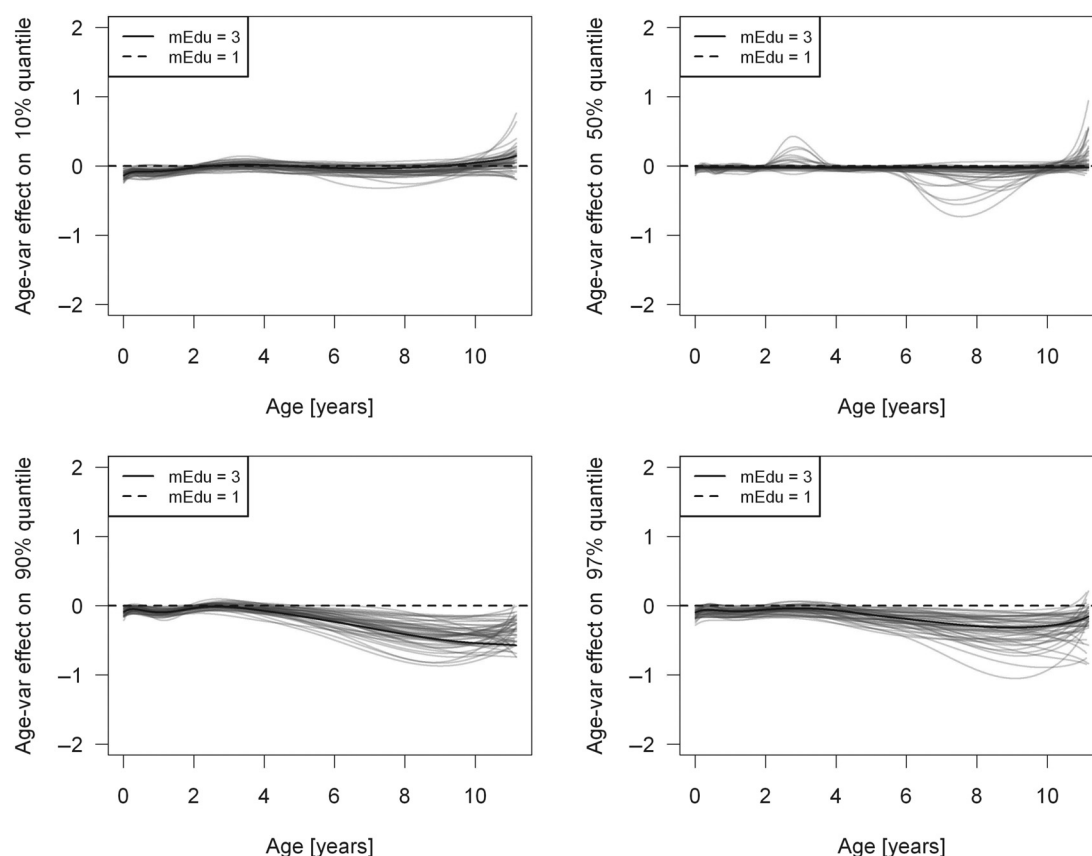


since the individual-specific intercepts estimated by STAQ models differ for the different quantiles. The upper quantile curve, in particular, seems to account better for the increasing skewness of the BMI distribution with increasing age.

The results of smooth nonlinear effects for covariates other than age are not shown but briefly described here. With regard to the effect of maternal BMI, the shape of all BMI quantile curves looks roughly similar. The effect increases with increasing maternal BMI and remains constant from maternal BMI values around  $30 \text{ kg/m}^2$ . The slope of the 97% BMI quantile is estimated to be larger than that of other quantiles. The effect of maternal BMI gain during pregnancy was estimated to be almost linear and slightly increasing throughout all models, which suggests that larger maternal BMI gains during pregnancy result in larger BMI values of children.

Regarding age-varying effects, Figure 8 exemplarily displays estimated age-varying effects for high compared to low maternal education. The effect of high maternal education is estimated to be almost zero for the BMI median and the 10% quantile. Yet, estimated upper BMI quantiles are smaller for children whose mothers have achieved a high school diploma (compared with children of mothers with “CSE or Hauptschule”). These effects are not present at birth and do not emerge before the age of around 5 years. Then, they show a continuing decrease until the age of 10 years. One should be cautious to attribute this effect on maternal education only, since maternal education is closely related to the socioeconomic status and can also be seen as a proxy of further life style factors.

The results for all other age-varying effects are not shown but just described. Concerning sex, conditional BMI quantiles for boys are estimated to be larger than those for girls, and the effect size is clearly varying with age. The results for study location suggest that upper BMI quantiles of children living in urban areas are



**Figure 8** Estimated age-varying effects for high maternal education ( $mEdu = 3$ ) compared to low maternal education ( $mEdu = 1$ ) resulting from STAQ models for all four quantile parameters. Age-varying effects are shown by solid black lines (full model) and gray lines (50 different bootstrap samples). Dashed line at zero corresponds to the reference category.



smaller than those of children from rural areas. Yet, this effect is not present for the median and the 10% quantile and not before the age of 7 years. Both age-varying and main effects of breastfeeding are estimated to be almost zero for all quantiles. Maternal smoking during pregnancy exerts a slightly positive effect between 3 and 6 years and no clear effect afterward, but the effect clearly varies with age. Age-varying effects of maternal education refer to the reference category of low maternal education ( $mEdu = 1$ , “CSE or Hauptschule”). The age-varying effect of high maternal education was already shown and discussed in Figure 8. Medium maternal education ( $mEdu = 2$ , “secondary school or Realschule”) does not show an age-varying effect.

Note that such age-varying effects as observed for sex, study location, maternal smoking, and education can only be detected with STAQ models, since they are only present for upper BMI quantiles.

## 5 Discussion

*Key findings.* The comparison of STAQ models with classical Gaussian AMMs suggested that STAQ models can better handle the age-specific BMI skewness and are thus more adequate than AMMs when the interest is directed toward overweight and obesity.

By using quantile regression, we obtained similar results as Beyerlein et al. [19, 20] with regard to time-constant risk factors. Apart from age, other risk factors also exert their effect in a different way on upper quantiles of the BMI distribution than on the mean. In our analysis here, we could additionally assess individual-specific BMI patterns during life-course and age-varying effects.

The results of the smooth age-varying effects of categorical covariates were particularly interesting. For several variables, quantile curves were estimated to be similar until a certain age period at which the age-varying effects emerged. For example, the effect of high maternal education did not become apparent before the age of around 5 years. Then, the effect size slowly increased until it was most pronounced at the age of 10 years. To the best of our knowledge, such age-varying effects have not yet been investigated in the obesity literature.

*Strengths and limitations of our analysis.* By applying STAQ regression for upper quantiles of the BMI distribution, it was possible to adequately and flexibly model the age-specific skewness of the BMI distribution while adjusting for other risk factors and individual-specific effects. Since we analyzed raw BMI values directly, our analysis did not require reference charts to construct a binary or Z-score response which are modeled in usual regression approaches for overweight and obesity.

We applied boosting for longitudinal quantile regression which is currently the only approach that can estimate individual-specific and smooth nonlinear effects as well as varying-coefficient terms in the same STAQ predictor. By including individual-specific effects, the model accounted for the temporal correlation between repeated measurements. Varying-coefficient terms enabled several age-varying effects of time-constant risk factors to be detected.

In addition, our life-course approach allowed to adequately model individual-specific BMI patterns, as illustrated in Figure 7. The figure shows individual-specific 90% BMI prediction intervals which were constructed in accordance with the approaches in Meinshausen [35] and Mayr et al. [36]. First, two separate models for the 5% and 95% BMI quantiles were estimated, and then, the individual-specific quantile predictions were taken as interval limits.

However, an inherent limitation of boosting is that it requires high computational effort. Furthermore, subsampling strategies have to be applied in order to obtain standard errors. This results in an even more computationally challenging and time-consuming fitting process – in particular when a large dataset with many individuals is analyzed, as was the case here.

*Implications for future research.* We believe that quantile boosting for STAQ models is a promising approach for further investigating risk factors for overweight and obesity, since analyzing upper quantiles of the BMI distribution is more adequate than analyzing the mean for this purpose. Furthermore, quantile regression does not incur information coarsening in the same way as usual binary regression approaches.

From an epidemiological point of view, further investigation of the age-varying effects could lead to additional insights and could, for example, explain differences between findings on the impact of risk factors in previous studies relying on different age populations. It would also be interesting to re-run the analysis with data from an additional time point (which is currently collected), since the BMI skewness becomes even more pronounced for older children.

Finally, it might have seemed obvious that quantile regression would be more adequate than Gaussian mean regression in our analysis. However, we would like to stress that the use of Gaussian AMMs for quantile modeling makes sense when the response distribution is homoscedastic and approximately Gaussian conditional on covariates. In a previous analysis of the LISA study [21], for example, observations were only available until the age of 6 years. AMMs could compete with the more complex quantile regression models, since the age-specific BMI skewness is not present until the age of 6 years (see Figure 1). In such cases, we would recommend to use the well-studied framework of AMMs for longitudinal quantile regression instead of applying more complex quantile regression strategies.

Although inspired by the specific research questions of our analysis, we emphasize that our modeling approach is general and can also be useful for other applications with similar data structures.

## Appendix: computational details

We used software from the statistical software R for estimation [37]. For fitting STAQ models, we used the function `gamboost` from package `mboost` [34] with option `family = QuantReg()`. Base learners for estimating individual-specific effects were specified with `brandom()`, while smooth nonlinear effects and age-varying effects of categorical covariates were estimated by using the function `bbs()`. In accordance with the description in Section 3.4, we used the following model calls to estimate STAQ models on the full LISA dataset:

```
staqFormula <- BMI ~ bols(Sex, Location, Nutri, Smoke, mEdu, df=5) +
  bbs(AgeC, df=5) + bbs(mBMIgainC, df=5) +
  bbs(mBMIC, df=5) + bbs(ageSexInt, df=5) +
  bbs(ageLocationInt, df=5) + bbs(ageNutriInt, df=5) +
  bbs(ageSmokeInt, df=5) + bbs(ageMedu2Int, df=5) +
  bbs(ageMedu3Int, df=5) + brandom(ID, df=5) +
  brandom(ID, by=AgeC, df=5)

staq90 <- gamboost(staqFormula, data=lisaLong,
  family=QuantReg(tau = 0.90),
  control=boost_control(mstop=5000, nu=0.4))
```

Note that `AgeC`, `mBMIgainC`, and `mBMIC` are mean-centered versions of the variables `Age`, `mDiffBMI`, and `mBMI`. All variables ending on `Int` are interaction variables between age and different levels of the categorical covariates which were defined to model age-varying effects.

For fitting AMMs, the function `gamm4` from package `gamm4` [30] can be specified as follows:

```
ammLisa <- gamm4(BMI ~ s(AgeC) + s(mBMIC) + s(mBMIgainC) +
  s(ageSexInt) + s(ageLocationInt) +
  s(ageNutriInt) + s(ageSmokeInt) +
  s(ageMedu2Int) + s(ageMedu3Int) +
  Sex + Location + Nutri + Smoke + mEdu +
  (1 + AgeC|cID), data=lisaLong)
```

Note that in our analysis we used the R package *amer* [29] for fitting AMMs. However, this package has been recently retired from the CRAN repository, since its scope of functionality is completely covered by the more sophisticated R package *gamm4* [30] which also provides a more detailed documentation and permanent maintenance.

**Acknowledgements:** We thank the LISA-plus study group for their work. The LISA-plus study was funded by grants of the German Federal Ministry for Education, Science, Research and Technology (Grant Nos. 01 EG 9705/2 and 01 EG 9732), and the six-year follow-up of the LISA-plus study was funded by the German Federal Ministry of Environment (IUF, FKS 20462296). The work of PR was supported by the Kompetenznetz Adipositas (Competence Network Obesity) funded by the Federal Ministry of Education and Research (FKZ: 01GI0826). Personnel and financial support by the Munich Center of Health Sciences (MC-Health), which contributed to this research, is gratefully acknowledged.

## References

1. Kosti R, Panagiotakos D. The epidemic of obesity in children and adolescents in the world. *Central Eur J Public Health* 2006;14:151–9.
2. Freedman D, Khan L, Serdula M, Dietz W, Srinivasan S, Berenson G. The relation of childhood BMI to adult adiposity: the Bogalusa heart study. *Pediatrics* 2005;115:22–7.
3. Sassi F, Devaux M, Cecchini M, Rusticelli E. The obesity epidemic: analysis of past and projected future trends in selected OECD countries. Technical Report No. 45, OECD Health Working Papers, 2009.
4. LISA-plus study group. Available at: <http://www.helmholtz-muenchen.de/epi1/forschung/arbeitsgruppen/arbeitsgruppe-1-umweltepidemiologie/projekte/lisa-plus/index.html>, 1998–2010. Accessed: 12 Jun 2013.
5. Haslam D., James W. Obesity. *Lancet* 2005;366:1197–209.
6. Reilly J, Armstrong J, Dorosty A, Emmett P, Ness A, Rogers I, et al. Early life risk factors for obesity in childhood: cohort study. *Br Med J* 2005;330:1357–63.
7. Fahrmeir L, Kneib T, Lang S. Penalized structured additive regression for space-time data: a Bayesian perspective. *Stat Sinica* 2004;14:731–61.
8. Koenker R. Quantile regression for longitudinal data. *J Multivariate Anal* 2004;91:74–89.
9. Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 2007;8:140–54.
10. Liu Y, Bottai M. Mixed-effects models for conditional quantiles with longitudinal data. *Int J Biostat* 2009;5:28.
11. Farcomeni A. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat Comput* 2012;22:141–52.
12. Reich B, Bondell H, Wang H. Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics* 2010;11:337–52.
13. Fenske N, Kneib T, Hothorn T. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *J Am Stat Assoc* 2011;106:494–510.
14. Koenker R. Quantile regression. *Economic Society Monographs*. Cambridge: Cambridge University Press, 2005.
15. Koenker R. Additive models for quantile regression: model selection and confidence band-aids. *Braz J Probability Stat* 2011;25:239–62.
16. Yue YR, Rue H. Bayesian inference for additive mixed quantile regression models. *Comput Stat Data Anal* 2011;55:84–96.
17. Lamerz A, Kuepper-Nybelen J, Wehle C, Bruning N, Trost-Brinkhues G, Brenner H, et al. Social class, parental education, and obesity prevalence in a study of six-year-old children in Germany. *Int J Obes* 2005;29:373–80.
18. Borghi E, de Onis M, Garza C, van den Broeck J, Frongillo E, Grummer-Strawn L, et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Stat Med* 2006;25:247–65.
19. Beyerlein A, Fahrmeir L, Mansmann U, Töschke A. Alternative regression models to assess increase in childhood BMI. *BMC Med Res Methodol* 2008;8:59.
20. Beyerlein A, Töschke A, von Kries R. Risk factors for childhood overweight: shift of the mean body mass index and shift of the upper percentiles: results from a cross-sectional study. *Int J Obes* 2010;34:642–8.
21. Fenske N, Fahrmeir L, Rzehak P, Höhle M. Detection of risk factors for obesity in early childhood with quantile regression methods for longitudinal data. Technical Report 38, Ludwig-Maximilians-Universität München, 2008. <http://epub.ub.uni-muenchen.de/6260/>, Accessed: 12 Jun 2013.

22. Breitfelder A, Wenig C, Wolfenstetter S, Rzehak P, Menn P, John J, et al. Relative weight-related costs of healthcare use by children – results from the two German birth cohorts, GINI-plus and LISA-plus. *Econ Hum Biol* 2011;9:302–15.
23. Rzehak P, Sausenthaler S, Koletzko S, Bauer CP, Schaaf B, von Berg A, et al. Period-specific growth, overweight and modification by breastfeeding in the GINI and LISA birth cohorts up to age 6 years. *Eur J Epidemiol* 2009;24:449–67.
24. Ong K, Ahmed M, Emmett P, Preece M, Dunger D. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *Br Med J* 2000;320:967–71.
25. Toschke A, Grote V, Koletzko B, von Kries R. Identifying children at high risk for overweight at school entry by weight gain during the first 2 years. *Arch Pediatr Adolesc Med* 2004;158:449–52.
26. Chen A, Pennell ML, Klebanoff MA, Rogan WJ, Longnecker MP. Maternal smoking during pregnancy in relation to child overweight: follow-up to age 8 years. *Int J Epidemiol* 2006;35:121–30.
27. Ruppert D, Wand M, Carroll R. Semiparametric regression. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2003.
28. Eilers P, Marx B. Flexible smoothing with B-splines and penalties. *Stat Sci* 1996;11:89–121.
29. Scheipl F. amer: additive mixed models with lme4, 2012. <http://CRAN.R-project.org/package=amer>, R package version 0.6.10.
30. Wood S. gamm4: generalized additive mixed models using mgcv and lme4, 2013. <http://CRAN.R-project.org/package=gamm4>, R package version 0.1–6.
31. Bühlmann P, Hothorn T. Boosting algorithms: regularization, prediction and model fitting (with discussion). *Stat Sci* 2007;22:477–505.
32. Hofner B, Hothorn T, Kneib T, Schmid M. A framework for unbiased model selection based on boosting. *J Comput Graphical Stat* 2011;20:956–71.
33. Hothorn T, Bühlmann P, Kneib T, Schmid M, Hofner B. Model-based boosting 2.0. *J Machine Learn Res* 2010;11:2109–13.
34. Hothorn T, Bühlmann P, Kneib T, Schmid M, Hofner B. mboost: model-based boosting, 2013. <http://CRAN.R-project.org/package=mboost>, R package version 2.2–2.
35. Meinshausen N. Quantile regression forests. *J Machine Learn Res* 2006;7:983–99.
36. Mayr A, Hothorn T, Fenske N. Prediction intervals for future BMI values of individual children – a non-parametric approach by quantile boosting. *BMC Med Res Methodol* 2012;12:6.
37. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. <http://www.R-project.org/>, ISBN 3-900051-07-0.